2004 NHIS Linked Mortality Files
Analytic Guidelines

February 2005

These analytic guidelines are the most current recommendations from the National Center for Health Statistics (NCHS) and the Data Linkage Program for use with the 2004 NHIS Linked Mortality Files. These guidelines will be updated on a periodic basis as more information is learned from the analyses of the NHIS Linked Mortality Files.

These guidelines address the following analytical topics:

I. Eligibility status

Only NHIS participants 18 years of age or older at time of interview were eligible for mortality linkage. Those 17 years or younger at the time of the NHIS interview were ineligible for matching with NDI records as were NHIS participants who had insufficient identifying data to create a NDI submission record. The eligibility status of NHIS participants for mortality follow-up is indicated by the variable *ELIGSTAT*. For mortality or survival analyses, users should keep only the records with a value of *ELIGSTAT = 1* and use the new sample weights, *WGT_NEW*, provided on the file. (See Detailed Notes for Selected Variables)

II. Vital status variable

The NHIS Linked Mortality files include NCHS's final determination of vital status (*MORSTAT*) that should be used for mortality outcomes and survival calculations. It is important for users to note that NDI match information, including date and cause of death, may be returned for NHIS participants who were ultimately determined not to be deceased by NCHS. The NDI match information is provided to allow researchers to redefine vital status based on alternative NDI matching criteria. Users should not determine mortality status through the cause of death information, which is found in the variables *CAUSEAVL* or from the ICD-9 and ICD-10 cause of death codes, rather they should first determine decedents using MORSTAT = 1 and analyze the corresponding cause of death information for these individuals.

Users should note that the variable *STATUS* provided on the file is the vital status assignment returned by the NDI, but does not reflect NCHS's final determination of vital status. (See Detailed Notes for Selected Variables)

III. Altering the criteria for assigning vita status

The 2004 Linked Mortality File includes the NCHS recommended vital status ascertainment (*MORSTAT*) for each eligible NHIS participant. For NHIS participants with a NDI record match, the file also includes the NCHS *CLASS* and *SCORE* for that record from which the determination of vital status was made. The class and score variables are included on the file so that users can alter the criteria for determining vital status and conduct their own sensitivity analyses. Users should refer to the NHIS Linked Mortality files matching methodology, Appendix C for examples of evaluation studies using different criteria to assign vital status for NHIS-NDI match records.

IV. New sample weights

For the 2004 NHIS Linked Mortality Files, the NHIS sample weights were adjusted to account for record non-responses for NHIS respondents who could not be linked due to insufficient identifying data. The new sample weights (*WGT_NEW)* are provided on the file. Users should note that the original sample weights (*WTFA*) also are provided on the files for the convenience of the researcher. Descriptive statistics for the original and adjusted adult sample weights are provided below for NHIS survey years 1986-2000.

Table 1: Description of adjusted survey weights (WGT_NEW) for adult, eligible NHIS respondents considered eligible for mortality follow-up: NHIS survey years 1986-2000

Survey Year	Total Eligible	Minimum	Maximum	Mean	Standard Deviation
1986	44659	0.0	21382.0	3878.6	1217.3
1987	88234	0.0	17838.0	1986.4	760.6
1988	88365	0.0	13765.0	2006.7	677.3
1989	84572	0.0	23704.0	2122.8	833.8
1990	86388	0.0	18803.0	2100.4	789.5
1991	86364	0.0	15152.0	2121.2	772.9
1992	91871	0.0	13055.0	2013.3	804.9
1993	79159	0.0	12038.0	2360.1	975.1
1994	83719	0.0	19638.0	2264.8	879.7
1995	72756	0.0	17086.0	2628.3	1080.0
1996	45274	0.0	23719.0	4260.3	1810.1
1997	73685	0.0	17077.0	2650.1	1290.9
1998	70663	0.0	17003.0	2792.1	1493.8
1999	69788	0.0	17305.0	2860.3	1568.8
2000	72123	0.0	19589.0	2796.5	1550.4

Table 2: Description of original unadjusted sample weight (WTFA) for adult eligible NHIS respondents: NHIS survey years 1986-2000

Survey Year	Total Eligible	Minimum	Maximum	Mean	Standard Deviation
1986	44659	345.0	21042.0	3878.6	1083.6
1987	88234	204.0	17540.0	1986.4	703.1
1988	88365	175.0	13486.0	2006.7	617.5
1989	84572	259.0	23388.0	2122.8	775.8
1990	86388	228.0	18688.0	2100.4	737.9
1991	86364	233.0	15067.0	2121.2	709.7
1992	91871	158.0	12716.0	2013.3	738.3
1993	79159	230.0	11694.0	2360.1	900.1
1994	83719	242.0	19244.0	2264.8	786.0
1995	72756	574.0	16636.0	2628.4	980.0
1996	45274	732.0	23220.0	4260.3	1621.3
1997	73685	0.0	16103.0	2650.1	1009.7
1998	70663	0.0	15532.0	2792.1	1111.3
1999	69788	0.0	15454.0	2860.3	1159.7
2000	72123	0.0	18114.0	2796.5	1149.7

V. Variance estimation

The data collected in the NHIS are obtained through a complex sample design involving stratification, clustering, and multistage sampling. In addition, the weighting factors are subject to adjustments. Because of this complex design and adjusted sampling weights, the direct application of standard statistical analysis methods for estimation and hypothesis testing may yield misleading results. Users of the NHIS Linked Mortality files should use computer software that provides the capability of variance estimation and hypothesis testing for complex sample designs.

For more information on design and estimation of the NHIS survey and variance estimation techniques, users should refer to the methods documentation located at http://www.cdc.gov/nchs/about/major/nhis/methods.htm. Users should note that variance estimation procedures differ for different NHIS survey design periods. For the 2004 Linked Mortality files, there are three different variance estimation approaches for the NHIS years 1986-1994, 1995-1996, and 1997-2000. Please refer to the following documentation:

Variance estimation 1987-1994
http://www.cdc.gov/nchs/about/major/nhis/sudaan.htm

Variance estimation 1995-1996
http://www.cdc.gov/nchs/data/nhis/96var.pdf

Variance estimation 1997-2005
http://www.cdc.gov/nchs/data/nhis/9705var.pdf

Although NCHS uses Taylor series linearization methods, in particular the SUDAAN software, for NHIS variance estimation, a summary of available software for the analysis of surveys with complex sample designs can be found at www.fas.harvard.edu/~stats/survey-soft/survey-soft.html. This site includes a selected list of review articles that is not specific to a single software package.

- Rao, Kumar and Roberts (1989), "Analysis of sample survey data involving categorical response variables: Methods and software," *Survey Methodology*, 15, 161-186.
- Rowe, Westlake, and Rose (1990), "Software for statistical and social survey analysis 1989-90," *Computational Statistics and Data Analysis*, 9, 317-340.
- Bausch and Bankhofer (1992), Statistical software packages for PCs - A market survey," *Statistischen Hefte*, 33, 283-306.
- Chell (1992), "A comparison of some software packages for survey analysis," in Westlake, Banks, Payne, and Orchard, *Survey and Statistical Computing*, Elsevier/North-Holland, 385-394.

VI. Combining Multiple NHIS Linked Mortality Files

Researchers may need to combine several survey years of the NHIS Linked Mortality files to get reliable estimates for rare health characteristics or mortality outcomes. However, combining data may require restructuring of design variables or adjustment of sample weights. In some cases, it may not be possible to combine surveys across certain years. Users should note that when combining data sets, it the data users' responsibility to examine the possible changes in the questionnaires as well as locations on the data files. For a complete discussion of sample design issues in the NHIS Linked Mortality files and combining multiple files, users should refer to Weighting and Sample Design Issues.

The sampling frame for the NHIS is redesigned approximately every ten years to better measure the changing U.S. population and to meet new survey objectives. The 2004 NHIS Linked Mortality Files include the NHIS years 1986-2000, which encompasses two sample design periods (1985-1994) and (1995-2004). In addition, although the 1995 and 1996 NHIS were collected using the same design scheme as for 1997-2004, the set of public design variables released with the 1995 and 1996 NHIS data files is different from that released with 1997-2000 NHIS. For this reason the 1995-1996 NHIS must be treated differently from 1997-2000 data.

Researchers who combine data across survey years under the same survey design are able to use the design variables (e.g. STRATUM and PSU) available on the NHIS data files. However, if the NHIS data sets are from different survey designs, different approaches for design based analysis must be implemented because there is no similarity or comparability in their sampling strata and PSUs. For a complete discussion and

examples of methods of combining NHIS Linked Mortality files across survey years as well as combining estimates from different NHIS survey designs, please refer to Weighting and Sample Design Issues.